PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/02, 21/04, 1/00, 17/00, C12Q 1/68, G01N 33/53	A1	(11) International Publication Number: WO 99/06427 (43) International Publication Date: 11 February 1999 (11.02.99)

(21) International Application Number: PCT/US98/16241

(22) International Filing Date: 4 August 1998 (04.08.98)

(30) Priority Data:
60/054,645 4 August 1997 (04.08.97) US

(71) Applicant: MILLENNIUM BIOTHERAPEUTICS, INC. [US/US]; 620 Memorial Drive, Cambridge, MA 02142 (US).

(72) Inventor: MCCARTHY, Sean, A.; 62 Commonwealth Avenue #4, Boston, MA 02116 (US).

(74) Agent: MEIKLEJOHN, Anita, L.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: TANGO-78, TANGO-79, AND TANGO-81 NUCLEIC ACID MOLECULES AND POLYPEPTIDES

(57) Abstract

The invention relates to Tango-78, Tango-79, and Tango-81 polypeptides, nucleic acid molecules encoding Tango-78, Tango-79, and Tango-81, and uses thereof.

B1

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TANGO-78, TANGO-79, AND TANGO-81 NUCLEIC ACID MOLECULES AND POLYPEPTIDES

Summary of the Invention

The invention relates to the discovery and characterization of the genes encoding Tango-78, Tango-79, and Tango-81.

The invention features isolated nucleic acid molecules encoding Tango-78, Tango-79, or Tango-81, the isolated nucleic acid molecules that encode polypeptides that are substantially identical to the Tango-78, Tango-79, or Tango-81 protein sequences described herein (SEQ ID NOS:2, 4, or 6) and isolated nucleic acid molecules which hybridize under stringent conditions to the protein coding portions of the Tango-78, Tango-79, or Tango-81 nucleic acid molecules described herein.

The invention also features a host cell which includes an isolated nucleic acid molecule encoding Tango-78, Tango-79, or Tango-81, a nucleic acid vector (e.g., an expression vector; a vector which includes a regulatory element; a vector which includes a regulatory element selected from the group consisting of the cytomegalovirus hCMV immediate early gene, the early promoter of SV40 adenovirus, the late promoter of SV40 adenovirus, the lac system, the trp system, the TAC system, the TRC system, the major operator and promoter regions of phage λ, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase, the promoters of acid phosphatase, and the promoters of the yeast α-mating factors; a vector which includes a regulatory element which directs tissue-specific expression; a vector which includes a reporter gene; a vector which includes a reporter gene selected from the group selected from the group consisting of β-lactamase, chloramphenicol acetyltransferase (CAT), adenosine deaminase (ADA), aminoglycoside phosphotransferase (neo^r,

G418^r), dihydrofolate reductase (DHFR), hygromycin-B-phosphotransferase (HPH), thymidine kinase (TK), lacZ (encoding β-galactosidase), and xanthine guanine phosphoribosyltransferase (XGPRT); a vector that is a plasmid; a vector that is a virus; a vector that is a retrovirus.

In another embodiment, the invention features a substantially pure Tango-78, Tango-79, or Tango-81 polypeptide (e.g., a Tango-78, Tango-79, or Tango-81 polypeptide that is soluble under physiological conditions; a Tango-78, Tango-79, or Tango-81 polypeptide which includes a signal sequence; a Tango-78 polypeptide that is at least 85%, 90%, 95%, or 100% identical to the amino acid sequence of SEQ ID NO:2; a Tango-79 polypeptide that is at least 85%, 90%, 95%, or 100% identical to the amino acid sequence of SEQ ID NO:4; and a Tango-81 polypeptide that is at least 85%, 90%, 95%, or 100% identical to the amino acid sequence of SEQ ID NO:6.

In other embodiments, the invention features a substantially pure polypeptide which includes a first portion and a second portion, the first portion including a Tango-78, Tango-79, or Tango-81 polypeptide and the second portion including a detectable marker.

The invention also features antibodies, e.g., monoclonal antibodies, that selectively binds to a polypeptide of the invention (Tango-78, Tango-79, or Tango-81).

The invention also features a pharmaceutical composition which includes a Tango-78, Tango-79, or Tango-81 polypeptide.

The invention also features a method for diagnosing a disorder associated with aberrant expression of Tango-78 the method including obtaining a biological sample from a patient and measuring Tango-78 expression in the biological sample, wherein increased or decreased

Tango-78 expression in the biological sample compared to a control indicates that the patient suffers from a disorder associated with aberrant expression of Tango-78.

The invention also features a method for diagnosing a disorder associated with aberrant expression of Tango-79, the method including obtaining a biological sample from a patient and measuring Tango-79 expression in the biological sample, wherein increased or decreased Tango-79 expression in the biological sample compared to a control indicates that the patient suffers from a disorder associated with aberrant expression of Tango-79.

The invention also features a method for diagnosing a disorder associated with aberrant expression of Tango-81, the method including obtaining a biological sample from a patient and measuring Tango-81 expression in the biological sample, wherein increased or decreased Tango-81 expression in the biological sample compared to a control indicates that the patient suffers from a disorder associated with aberrant expression of Tango-81.

The invention encompasses isolated nucleic acid molecules encoding Tango-78, Tango-79, or Tango-81 or a polypeptide fragment thereof; vectors containing these nucleic acid molecules; cells harboring recombinant DNA encoding Tango-78, Tango-79, or Tango-81; fusion proteins which include Tango-78, Tango-79, or Tango-81; transgenic animals which express Tango-78, Tango-79, or Tango-81; recombinant knock-out animals which fail to express Tango-78, Tango-79, or Tango-81.

The invention encompasses nucleic acids that have a sequence that is substantially identical to the nucleic acid sequence of Tango-78, Tango-79, or Tango-81. A nucleic acid sequence which is substantially identical to a given reference nucleic acid sequence is hereby defined as a nucleic acid having a sequence that has at least 85%, preferably 90%, and more preferably 95%, 98%, 99% or

more identity to the sequence of the given reference nucleic acid sequence, e.g., the nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5.

The invention encompasses polypeptides that have a sequence that is substantially identical to the amino acid sequence of Tango-78, Tango-79, or Tango-81. A polypeptide which is "substantially identical" to a given reference polypeptide is a polypeptide having a sequence that has at least 85%, preferably 90%, and more preferably 95%, 98%, 99% or more identity to the sequence of the given reference polypeptide sequence, e.g., the amino sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6.

The nucleic acid molecules of the invention can be inserted into vectors, as described below, which will facilitate expression of the insert. The nucleic acid molecules and the polypeptides they encode can be used directly as diagnostic or therapeutic agents, or (in the case of a polypeptide) can be used to generate antibodies that, in turn, are therapeutically useful. Accordingly, expression vectors containing the nucleic acid molecules of the invention, cells transfected with these vectors, the polypeptides expressed, and antibodies generated (against either the entire polypeptide or an antigenic fragment thereof) are among the preferred embodiments.

A transformed cell is any cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a nucleic acid encoding a polypeptide of the invention (e.g., a Tango-78, Tango-79, or Tango-81 polypeptide).

An isolated nucleic acid molecule is a nucleic acid molecule that is separated from the 5′ and 3′ coding sequences with which it is immediately contiguous in the naturally occurring genome of an organism. Isolated nucleic acid molecules include nucleic acid molecule

which are not naturally occurring, e.g., nucleic acid molecules created by recombinant DNA techniques.

Nucleic acid molecules include both RNA and DNA, including cDNA, genomic DNA, and synthetic (e.g., chemically synthesized) DNA. Where single-stranded, the nucleic acid molecule may be a sense strand or an antisense strand.

The invention also encompasses nucleic acid molecules that hybridize, preferably under stringent conditions, to a nucleic acid molecule encoding a Tango-78, Tango-79, or Tango-81 polypeptide (e.g., the polypeptide encoding portions of SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5). Preferably the hybridizing nucleic acid molecule consists of 400, more preferably 200 nucleotides. Preferred hybridizing nucleic acid molecules have a biological activity possessed by Tango-78, Tango-79, or Tango-81.

The invention also features substantially pure or isolated Tango-78, Tango-79, or Tango-81 polypeptides, including those that correspond to various functional domains of Tango-78, Tango-79, or Tango-81, or fragments thereof.

The polypeptides of the invention can be prepared by recombinant gene expression, chemically synthesized, or purified from tissues in which they are naturally expressed using standard biochemical methods of purification.

Also included in the invention are functional polypeptides, which possess one or more of the biological functions or activities of Tango-78, Tango-79, or Tango-81. These functions include the ability to bind some or all of the proteins which normally bind to Tango-78, Tango-79, or Tango-81. A functional polypeptide is also considered within the scope of the invention if it serves as an antigen for production of antibodies that

specifically bind to Tango-78, Tango-79, or Tango-81. In many cases, functional polypeptides retain one or more domains present in the naturally-occurring form of the polypeptide.

The functional polypeptides may contain a primary amino acid sequence that has been modified from those disclosed herein. Preferably these modifications consist of conservative amino acid substitutions, as described herein.

The terms "protein" and "polypeptide" are used herein interchangably to describe any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation). Thus, the term "Tango-78, Tango-79, or Tango-81 polypeptide" includes: full-length, naturally occurring Tango-78, Tango-79, or Tango-81 protein; recombinantly or synthetically produced polypeptide that corresponds to a full-length naturally occurring Tango-78, Tango-79, or Tango-81; or particular domains or portions of the naturally occurring protein. The term also encompasses mature Tango-78, Tango-79, or Tango-81 which has an added amino-terminal methionine (useful for expression in prokaryotic cells).

The term "purified" as used herein refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized.

Polypeptides or other compounds of interest are said to be "substantially pure" when they are within preparations that are at least 60% by weight (dry weight) the compound of interest. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight the compound of

interest. Purity can be measured by any appropriate standard method, for example, by column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

Where a particular polypeptide or nucleic acid molecule is said to have a specific percent identity to a reference polypeptide or nucleic acid molecule of a defined length, the percent identity is relative to the reference polypeptide or nucleic acid molecule. Thus, a peptide that is 50% identical to a reference polypeptide that is 100 amino acids long can be a 50 amino acid polypeptide that is completely identical to a 50 amino acid long portion of the reference polypeptide. It might also be a 100 amino acid long polypeptide which is 50% identical to the reference polypeptide over its entire length. Of course, many other polypeptides will meet the same criteria. The same rule applies for nucleic acid molecules.

For polypeptides, the length of the reference polypeptide sequence will generally be at least 16 amino acids, preferably at least 20 amino acids, more preferably at least 25 amino acids, and most preferably 35 amino acids, 50 amino acids, or 100 amino acids. For nucleic acids, the length of the reference nucleic acid sequence will generally be at least 50 nucleotides, preferably at least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably 100 nucleotides or 300 nucleotides.

In the case of polypeptide sequences which are less than 100% identical to a reference sequence, the non-identical positions are preferably, but not necessarily, conservative substitutions for the reference sequence. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine, and leucine; aspartic acid and glutamic acid; asparagine and glutamine; serine and

threonine; lysine and arginine; and phenylalanine and tyrosine.

Sequence identity can be measured using sequence analysis software (for example, the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705), with the default parameters as specified therein.

The invention also features antibodies, e.g., monoclonal, polyclonal, and engineered antibodies, which specifically bind Tango-78, Tango-79, or Tango-81. By "specifically binds" is meant an antibody that recognizes and binds to a particular antigen, e.g., a Tango-78, Tango-79, or Tango-81 polypeptide of the invention, but which does not substantially recognize or bind to other molecules in a sample, e.g., a biological sample, which includes the polypeptide.

The invention also features antagonists and agonists of Tango-78, Tango-79, or Tango-81 that can inhibit or enhance, respectively, one or more of the biological activities of Tango-78, Tango-79, or Tango-81. Suitable antagonists can include small molecules (i.e., molecules with a molecular weight below about 500); large molecules (i.e., molecules with a molecular weight above about 500), antibodies that bind and "neutralize" Tango-78, Tango-79, or Tango-81 (as described below); polypeptides which compete with a native form of Tango-78, Tango-79, or Tango-81 for binding to a functional binding partner of the native protein; and nucleic acid molecules that interfere with transcription of Tango-78, Tango-79, or Tango-81 (for example, antisense nucleic acid molecules and ribozymes). Agonists of Tango-78, Tango-79, or Tango-81 also include small and large molecules, and antibodies other than neutralizing antibodies.

The invention also features molecules which can increase or decrease the expression of Tango-78, Tango-79, or Tango-81 (e.g., by influencing transcription or translation). Small molecules (i.e., molecules with a molecular weight below about 500), large molecules (i.e., molecules with a molecular weight above about 500), and nucleic acid molecules that can be used to inhibit the expression of Tango-78, Tango-79, or Tango-81 (for example, antisense and ribozyme molecules) or to enhance their expression (for example, molecules that bind to a Tango-78, Tango-79, or Tango-81 transcription regulatory sequence and increase transcription.

In addition, the invention features substantially pure polypeptides that functionally interact with Tango-78, Tango-79, or Tango-81 and the nucleic acid molecules that encode them.

The invention encompasses methods for treating disorders associated with aberrant expression or activity of a protein of the invention (i.e., Tango-78, Tango-79, or Tango-81). Thus, the invention includes methods for treating disorders associated with excessive expression or activity of the protein. Such methods entail administering a compound which decreases the expression of the protein. The invention also includes methods for treating disorders associated with insufficient expression or activity of a protein of the invention. These methods entail administering a compound which increases the expression or activity of the protein.

The invention also features methods for detecting a protein of the invention. Such methods include: obtaining a biological sample; contacting the sample with an antibody that specifically binds to the protein under conditions which permit specific binding; and detecting any antibody-protein complexes formed.

In addition, the present invention encompasses methods and compositions for the diagnostic evaluation, typing, and prognosis of disorders associated with inappropriate expression or activity of Tango-78, Tango-79, or Tango-81. For example, the nucleic acid molecules of the invention can be used as diagnostic hybridization probes to detect, for example, inappropriate expression of Tango-78, Tango-79, or Tango-81 or mutations in the Tango-78, Tango-79, or Tango-81 gene. Such methods may be used to classify cells by the level of Tango-78, Tango-79, or Tango-81 expression.

Thus, the invention features a method for diagnosing a disorder associated with aberrant activity of a protein of the invention, the method including obtaining a biological sample from a patient and measuring the activity of the protein in the biological sample, wherein increased or decreased activity in the biological sample compared to a control indicates that the patient suffers from a disorder associated with aberrant activity of the protein.

The nucleic acid molecules can be used as primers for diagnostic PCR analysis for the identification of gene mutations, allelic variations and regulatory defects in the Tango-78, Tango-79, or Tango-81 gene. The present invention further provides for diagnostic kits for the practice of such methods.

The invention features methods of identifying compounds that modulate the expression or activity of a protein of the invention by assessing the expression or activity of the protein in the presence and absence of a selected compound. A difference in the level of expression or activity of the protein in the presence and absence of the selected compound indicates that the selected compound is capable of modulating expression or activity of the protein. Expression can be assessed

either at the level of gene expression (e.g., by measuring mRNA) or protein expression by techniques that are well known to skilled artisans.

The preferred methods and materials are described below in examples which are meant to illustrate, not limit, the invention. Skilled artisans will recognize methods and materials that are similar or equivalent to those described herein, and that can be used in the practice or testing of the present invention.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described herein. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting.

Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

Brief Description of the Drawing

Figure 1 is a depiction of the nucleic acid sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) of Tango-78.

Figure 2 is a depiction of the nucleic sequence (SEQ ID NO:3) and deduced amino acid sequence (SEQ ID NO:4) of Tango-79.

Figure 3 is a depiction of the nucleic acid sequence (SEQ ID NO:5) and deduced amino acid sequence (SEQ ID NO:6) of Tango-81.

Figure 4 is an alignment of the amion acid sequence of Tango-78 and the amion acid sequence of murine nodal protein.

Figure 5 is an alignment between the amino acid sequence of Tango-79 and D45913 (Leucine rich repeat protein).

Figure 6 is a depiction of the results of Northern blot analysis of Tango-81 expression.

Detailed Description

Tango-78, Tango-79, and Tango-81 Nucleic Acid 27 Molecules

The Tango-78, Tango-79, and Tango-81 nucleic acid molecules of the invention can be cDNA, genomic DNA, synthetic DNA, or RNA, and can be double-stranded or single-stranded (*i.e.*, either a sense or an antisense strand). Fragments of these molecules are also considered within the scope of the invention, and can be produced, for example, by the polymerase chain reaction (PCR) or generated by treatment with one or more restriction endonucleases. A ribonucleic acid (RNA) molecule can be produced by *in vitro* transcription.

The nucleic acid molecules of the invention can contain naturally occurring sequences, or sequences that differ from those that occur naturally, but, due to the degeneracy of the genetic code, encode the same polypeptide. In addition, these nucleic acid molecules are not limited to sequences that only encode polypeptides, and thus, can include some or all of the non-coding sequences that lie upstream or downstream from a coding sequence.

The nucleic acid molecules of the invention can be synthesized (for example, by phosphoramidite-based synthesis) or obtained from a biological cell, such as the cell of a mammal. Thus, the nucleic acids can be those of a human, mouse, rat, guinea pig, cow, sheep, horse, pig, rabbit, monkey, dog, or cat. Combinations or modifications of the nucleotides within these types of nucleic acids are also encompassed.

In addition, the isolated nucleic acid molecules of the invention encompass fragments that are not found as such in the natural state. Thus, the invention encompasses recombinant molecules, such as those in which a nucleic acid molecule (for example, an isolated nucleic acid molecule encoding Tango-78, Tango-79, or Tango-81) is incorporated into a vector (for example, a plasmid or viral vector) or into the genome of a heterologous cell (or the genome of a homologous cell, at a position other than the natural chromosomal location). Recombinant nucleic acid molecules and uses therefor are discussed further below.

In the event the nucleic acid molecules of the invention encode or act as antisense molecules, they can be used for example, to regulate translation of Tango-78, Tango-79, or Tango-81 mRNA.

The invention also encompasses nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule encoding a Tango-78, Tango-79, or Tango-81 polypeptide (e.g., the protein encoding portion of SEQ ID NO:1, SEQ ID:3, or SEQ ID NO:5). The cDNA sequences described herein can be used to identify these nucleic acids, which include, for example, nucleic acids that encode homologous polypeptides in other species, and splice variants of the Tango-78, Tango-79, or Tango-81 gene in humans or other mammals. Accordingly, the invention features methods of detecting and isolating

these nucleic acid molecules. Using these methods, a sample (for example, a nucleic acid library, such as a cDNA or genomic library) is contacted (or "screened") with a Tango-78, Tango-79, or Tango-81-specific probe. The probe will selectively hybridize to nucleic acids encoding related polypeptides (or to complementary sequences thereof). The probe, which can contain at least 25 (for example, 25, 50, 100, or 200 nucleotides) can be produced using any of several standard methods (see, for example, Ausubel et al.,"Current Protocols in Molecular Biology, Vol. I," Green Publishing Associates, Inc., and John Wiley & Sons, Inc., NY, 1989). For example, the probe can be generated using PCR amplification methods in which oligonucleotide primers are used to amplify a Tango-78, Tango-79, or Tango-81-specific nucleic acid sequence that can be used as a probe to screen a nucleic acid library and thereby detect nucleic acid molecules (within the library) that hybridize to the probe.

One single-stranded nucleic acid is said to hybridize to another if a duplex forms between them. This occurs when one nucleic acid contains a sequence that is the reverse and complement of the other (this same arrangement gives rise to the natural interaction between the sense and antisense strands of DNA in the genome and underlies the configuration of the "double helix"). Complete complementarity between the hybridizing regions is not required in order for a duplex to form; it is only necessary that the number of paired bases is sufficient to maintain the duplex under the hybridization conditions used.

Typically, hybridization conditions are of low to moderate stringency. These conditions favor specific interactions between completely complementary sequences, but allow some non-specific interaction between less than

perfectly matched sequences to occur as well. After hybridization, the nucleic acids can be "washed" under moderate or high conditions of stringency to dissociate duplexes that are bound together by some non-specific interaction (the nucleic acids that form these duplexes are thus not completely complementary).

As is known in the art, the optimal conditions for washing are determined empirically, often by gradually increasing the stringency. The parameters that can be changed to affect stringency include, primarily, temperature and salt concentration. In general, the lower the salt concentration and the higher the temperature, the higher the stringency. Washing can be initiated at a low temperature (for example, room temperature) using a solution containing a salt concentration that is equivalent to or lower than that of the hybridization solution. Subsequent washing can be carried out using progressively warmer solutions having the same salt concentration. As alternatives, the salt concentration can be lowered and the temperature maintained in the washing step, or the salt concentration can be lowered and the temperature increased. Additional parameters can also be altered. For example, use of a destabilizing agent, such as formamide, alters the stringency conditions.

In reactions where nucleic acids are hybridized, the conditions used to achieve a given level of stringency will vary. There is not one set of conditions, for example, that will allow duplexes to form between all nucleic acids that are 85% identical to one another; hybridization also depends on unique features of each nucleic acid. The length of the sequence, the composition of the sequence (for example, the content of purine-like nucleotides versus the content of pyrimidine-like nucleotides) and the type of nucleic acid (for

example, DNA or RNA) affect hybridization. An additional consideration is whether one of the nucleic acids is immobilized (for example, on a filter).

An example of a progression from lower to higher stringency conditions is the following, where the salt content is given as the relative abundance of SSC (a salt solution containing sodium chloride and sodium citrate; 2X SSC is 10-fold more concentrated than 0.2X SSC). Nucleic acids are hybridized at 42°C in 2X SSC/0.1% SDS (sodium dodecylsulfate; a detergent) and then washed in 0.2X SSC/0.1% SDS at room temperature (for conditions of low stringency); 0.2X SSC/0.1% SDS at 42°C (for conditions of moderate stringency); and 0.1X SSC at 68°C (for conditions of high stringency). Washing can be carried out using only one of the conditions given, or each of the conditions can be used (for example, washing for 10-15 minutes each in the order listed above). Any or all of the washes can be repeated. As mentioned above, optimal conditions will vary and can be determined empirically.

A second set of conditions that are considered "stringent conditions" are those in which hybridization is carried out at 50°C in Church buffer (7% SDS, 0.5% $NaHPO_4$, 1 M EDTA, 1% BSA) and washing is carried out at 50°C in 2X SSC.

Once detected, the nucleic acid molecules can be isolated by any of a number of standard techniques (see, for example, Sambrook et al., "Molecular Cloning, A Laboratory Manual," 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

The invention also encompasses: (a) expression vectors that contain any of the foregoing Tango-78, Tango-79, and Tango-81-related coding sequences and/or their complements (that is, "antisense" sequence); (b) expression vectors that contain any of the foregoing

Tango-78, Tango-79, or Tango-81-related coding sequences operatively associated with a regulatory element (examples of which are given below) that directs the expression of the coding sequences; (c) expression vectors containing, in addition to sequences encoding a Tango-78, Tango-79, or Tango-81 polypeptide, nucleic acid sequences that are unrelated to nucleic acid sequences encoding Tango-78, Tango-79, or Tango-81, such as molecules encoding a reporter or marker; and (d) genetically engineered host cells that contain any of the foregoing expression vectors and thereby express the nucleic acid molecules of the invention in the host cell.

Recombinant nucleic acid molecules can contain a sequence encoding a soluble Tango-78, Tango-79, or Tango-81 polypeptide; mature Tango-78, Tango-79, or Tango-81; or Tango-78, Tango-79, or Tango-81 having an added or endogenous signal sequence. A full length Tango-78, Tango-79, or Tango-81 polypeptide; a domain of Tango-78, Tango-79, or Tango-81; or a fragment thereof may be fused to additional polypeptides, as described below. Similarly, the nucleic acid molecules of the invention can encode the mature form of Tango-78, Tango-79, or Tango-81 or a form that encodes a polypeptide which facilitates secretion. In the latter instance, the polypeptide is typically referred to as a proprotein, which can be converted into an active form by removal of the signal sequence, for example, within the host cell. Proproteins can be converted into the active form of the protein by removal of the inactivating sequence.

The regulatory elements referred to above include, but are not limited to, inducible and non-inducible promoters, enhancers, operators and other elements, which are known to those skilled in the art, and which drive or otherwise regulate gene expression. Such regulatory elements include but are not limited to the

cytomegalovirus hCMV immediate early gene, the early or late promoters of SV40 adenovirus, the lac system, the trp system, the TAC system, the TRC system, the major operator and promoter regions of phage A, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase, the promoters of acid phosphatase, and the promoters of the yeast α-mating factors.

Similarly, the nucleic acid can form part of a hybrid gene encoding additional polypeptide sequences, for example, sequences that function as a marker or reporter. Examples of marker or reporter genes include β-lactamase, chloramphenicol acetyltransferase (CAT), adenosine deaminase (ADA), aminoglycoside phosphotransferase (neo^r, $G418^r$), dihydrofolate reductase (DHFR), hygromycin-B-phosphotransferase (HPH), thymidine kinase (TK), lacZ (encoding β-galactosidase), and xanthine guanine phosphoribosyltransferase (XGPRT). As with many of the standard procedures associated with the practice of the invention, skilled artisans will be aware of additional useful reagents, for example, of additional sequences that can serve the function of a marker or reporter. Generally, the hybrid polypeptide will include a first portion and a second portion; the first portion being a Tango-78, Tango-79, or Tango-81 polypeptide and the second portion being, for example, the reporter described above or an immunoglobulin constant region.

The expression systems that may be used for purposes of the invention include, but are not limited to, microorganisms such as bacteria (for example, *E. coli* and *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA, or cosmid DNA expression vectors containing the nucleic acid molecules of the invention; yeast (for example, *Saccharomyces* and *Pichia*) transformed with recombinant yeast expression vectors

containing the nucleic acid molecules of the invention; insect cell systems infected with recombinant virus expression vectors (for example, baculovirus) containing the nucleic acid molecules of the invention; plant cell systems infected with recombinant virus expression vectors (for example, cauliflower mosaic virus (CaMV) and tobacco mosaic virus (TMV)) or transformed with recombinant plasmid expression vectors (for example, Ti plasmid) containing Tango-78, Tango-79, or Tango-81 nucleotide sequences; or mammalian cell systems (for example, COS, CHO, BHK, 293, VERO, HeLa, MDCK, WI38, and NIH 3T3 cells) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (for example, the metallothionein promoter) or from mammalian viruses (for example, the adenovirus late promoter and the vaccinia virus 7.5K promoter).

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the gene product being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of pharmaceutical compositions containing Tango-78, Tango-79, or Tango-81 polypeptides or for raising antibodies to those polypeptides, vectors that are capable of directing the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited to, the *E. coli* expression vector pUR278 (Ruther et al., *EMBO J.* 2:1791, 1983), in which the coding sequence of the insert may be ligated individually into the vector in frame with the lacZ coding region so that a fusion protein is produced; pIN vectors (Inouye and Inouye, *Nucleic Acids Res.* 13:3101-3109, 1985; Van Heeke and Schuster, *J. Biol. Chem.* 264:5503-5509, 1989); and the like. pGEX vectors may

also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

In an insect system, *Autographa californica* nuclear polyhidrosis virus (AcNPV) can be used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The coding sequence of the insert may be cloned individually into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of the coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (i.e., virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed. (for example, see Smith et al., *J. Virol.* 46:584, 1983; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the nucleic acid molecule of the invention may be ligated to an adenovirus transcription/translation control complex, for example, the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (for example, region E1 or E3) will result in a

recombinant virus that is viable and capable of expressing a Tango-78, Tango-79, or Tango-81 gene product in infected hosts (for example, see Logan and Shenk, *Proc. Natl. Acad. Sci. USA* 81:3655-3659, 1984). Specific initiation signals may also be required for efficient translation of inserted nucleic acid molecules. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire gene or cDNA, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only a portion of the coding sequence is inserted, exogenous translational control signals, including, perhaps, the ATG initiation codon, must be provided. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bittner et al., *Methods in Enzymol.* 153:516-544, 1987).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (for example, glycosylation) and processing (for example, cleavage) of protein products may be important for the function of the protein. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and

processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. The mammalian cell types listed above are among those that could serve as suitable host cells.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the Tango-78, Tango-79, or Tango-81 sequences described above may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (for example, promoter, enhancer sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method can advantageously be used to engineer cell lines which express Tango-78, Tango-79, or Tango-81. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that affect the endogenous activity of the gene product.

A number of selection systems can be used. For example, the herpes simplex virus thymidine kinase (Wigler, et al., *Cell* 11:223, 1977), hypoxanthine-guanine phosphoribosyltransferase (Szybalska and Szybalski, *Proc. Natl. Acad. Sci. USA* 48:2026, 1962), and adenine phosphoribosyltransferase (Lowy, et al., *Cell* 22:817,

1980) genes can be employed in tk^-, $hgprt^-$ or $aprt^-$ cells, respectively. Also, anti-metabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler et al., *Proc. Natl. Acad. Sci. USA* 77:3567, 1980; O'Hare et al., *Proc. Natl. Acad. Sci. USA* 78:1527, 1981); gpt, which confers resistance to mycophenolic acid (Mulligan and Berg, *Proc. Natl. Acad. Sci. USA* 78:2072, 1981); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin et al., *J. Mol. Biol.* 150:1, 1981); and hygro, which confers resistance to hygromycin (Santerre et al., *Gene* 30:147, 1984).

The nucleic acid molecules of the invention are useful for diagnosis of disorders associated with aberrant expression of Tango-78, Tango-79, or Tango-81. Tango-78, Tango-79, and Tango-81 nucleic acid molecules are also useful in genetic mapping and chromosome identification.

Tango-78, Tango-79, and Tango-81 Polypeptides

The Tango-78, Tango-79, and Tango-81 polypeptides described herein are those encoded by any of the nucleic acid molecules described above and include Tango-78, Tango-79, and Tango-81 fragments, mutants, truncated forms, and fusion proteins. These polypeptides can be prepared for a variety of uses, including but not limited to the generation of antibodies, as reagents in diagnostic assays, for the identification of other cellular gene products or compounds that can modulate the activity or expression of Tango-78, Tango-79, or Tango-81, and as pharmaceutical reagents useful for the treatment of disorders associated with aberrant expression or activity of Tango-78, Tango-79, or Tango-81.

Preferred polypeptides are substantially pure Tango-78, Tango-79, and Tango-81 polypeptides, including

those that correspond to the polypeptide with an intact signal sequence, and the secreted form of the polypeptide.

The invention also encompasses polypeptides that are functionally equivalent to Tango-78, Tango-79, or Tango-81. These polypeptides are equivalent to Tango-78, Tango-79, or Tango-81 in that they are capable of carrying out one or more of the functions of Tango-78, Tango-79, or Tango-81 in a biological system. Preferred Tango-78, Tango-79, or Tango-81 polypeptides have 20%, 40%, 50%, 75%, 80%, or even 90% of one or more of the biological activities of the full-length, mature human form of Tango-78, Tango-79, and Tango-81. Such comparisons are generally based on an assay of biological activity in which equal concentrations of the polypeptides are used and compared. The comparison can also be based on the amount of the polypeptide required to reach 50% of the maximal stimulation obtainable.

Functionally equivalent proteins can be those, for example, that contain additional or substituted amino acid residues. Substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. Amino acids that are typically considered to provide a conservative substitution for one another are specified in the summary of the invention.

Polypeptides that are functionally equivalent to Tango-78, Tango-79, or Tango-81 can be made using random mutagenesis techniques well known to those skilled in the art. It is more likely, however, that such polypeptides will be generated by site-directed mutagenesis (again using techniques well known to those skilled in the art). These polypeptides may have increased functionality or decreased functionality.

To design functionally equivalent polypeptides, it is useful to distinguish between conserved positions and variable positions. This can be done by aligning the amino acid sequence of a protein of the invention from one species with its homolog from another species. Skilled artisans will recognize that conserved amino acid residues are more likely to be necessary for preservation of function. Thus, it is preferable that conserved residues are not altered.

Mutations within the coding sequence of nucleic acid molecules of the invention can be made to generate variant genes that are better suited for expression in a selected host cell. For example, N-linked glycosylation sites can be altered or eliminated to achieve, for example, expression of a homogeneous product that is more easily recovered and purified from yeast hosts which are known to hyperglycosylate N-linked sites. To this end, a variety of amino acid substitutions at one or both of the first or third amino acid positions of any one or more of the glycosylation recognition sequences which occur, and/or an amino acid deletion at the second position of any one or more of such recognition sequences, will prevent glycosylation at the modified tripeptide sequence (see, for example, Miyajima et al., *EMBO J.* 5:1193, 1986).

The polypeptides of the invention can be expressed fused to another polypeptide, for example, a marker polypeptide or fusion partner. For example, the polypeptide can be fused to a hexa-histidine tag to facilitate purification of bacterially expressed protein or a hemagglutinin tag to facilitate purification of protein expressed in eukaryotic cells.

A fusion protein may be readily purified by utilizing an antibody specific for the fusion protein being expressed. For example, a system described by

Janknecht et al. allows for the ready purification of non-denatured fusion proteins expressed in human cell lines (*Proc. Natl. Acad. Sci. USA* 88: 8972-8976, 1991). In this system, the gene of interest is subcloned into a vaccinia recombination plasmid such that the gene's open reading frame is translationally fused to an amino-terminal tag consisting of six histidine residues. Extracts from cells infected with recombinant vaccinia virus are loaded onto Ni^{2+}·nitriloacetic acid-agarose columns and histidine-tagged proteins are selectively eluted with imidazole-containing buffers.

The polypeptides of the invention can be chemically synthesized (for example, see Creighton, "Proteins: Structures and Molecular Principles," W.H. Freeman & Co., NY, 1983), or, perhaps more advantageously, produced by recombinant DNA technology as described herein. For additional guidance, skilled artisans may consult Ausubel et al. (*supra*), Sambrook et al. ("Molecular Cloning, A Laboratory Manual," Cold Spring Harbor Press, Cold Spring Harbor, NY, 1989), and, particularly for examples of chemical synthesis Gait, M.J. Ed. ("Oligonucleotide Synthesis," IRL Press, Oxford, 1984).

The invention also features polypeptides that interact with Tango-78, Tango-79, or Tango-81 (and the genes that encode them) and thereby alter the function of Tango-78, Tango-79, or Tango-81. Interacting polypeptides can be identified using methods known to those skilled in the art. One suitable method is the "two-hybrid system," which detects protein interactions *in vivo* (Chien et al., *Proc. Natl. Acad. Sci. USA*, 88:9578, 1991). A kit for practicing this method is available from Clontech (Palo Alto, CA).

Transgenic animals

Tango-78, Tango-79, and Tango-81 polypeptides can also be expressed in transgenic animals. These animals represent a model system for the study of disorders that are caused by or exacerbated by overexpression or underexpression of Tango-78, Tango-79, or Tango-81, and for the development of therapeutic agents that modulate the expression or activity of Tango-78, Tango-79, or Tango-81.

Transgenic animals can be farm animals (pigs, goats, sheep, cows, horses, rabbits, and the like) rodents (such as rats, guinea pigs, and mice), non-human primates (for example, baboons, monkeys, and chimpanzees), and domestic animals (for example, dogs and cats). Transgenic mice are especially preferred.

Any technique known in the art can be used to introduce a Tango-78, Tango-79, or Tango-81 transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (U.S. Pat. No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82:6148, 1985); gene targeting into embryonic stem cells (Thompson et al., *Cell* 56:313, 1989); and electroporation of embryos (Lo, *Mol. Cell. Biol.* 3:1803, 1983).

The present invention provides for transgenic animals that carry a Tango-78, Tango-79, or Tango-81 transgene in all their cells, as well as animals that carry a transgene in some, but not all of their cells. That is, the invention provides for mosaic animals. The transgene can be integrated as a single transgene or in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene can also be selectively introduced into and activated in a particular cell type (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89:6232, 1992).

The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

When it is desired that the Tango-78, Tango-79, or Tango-81 transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be used, vectors containing some nucleotide sequences homologous to an endogenous Tango-78, Tango-79, or Tango-81 gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene also can be selectively introduced into a particular cell type, thus inactivating the endogenous Tango-78, Tango-79, or Tango-81 gene in only that cell type (Gu et al., *Science* 265:103, 1984). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. These techniques are useful for preparing "knock outs" lacking a functional gene.

Once transgenic animals have been generated, the expression of the recombinant Tango-78, Tango-79, or Tango-81 gene can be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to determine whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR. Biological samples can also be evaluated

immunocytochemically using antibodies specific for the Tango-78, Tango-79, or Tango-81 transgene product.

For a review of techniques that can be used to generate and assess transgenic animals, skilled artisans can consult Gordon (*Intl. Rev. Cytol*. 115:171-229, 1989), and may obtain additional guidance from, for example: Hogan et al. "Manipulating the Mouse Embryo" (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1986; Krimpenfort et al., *Bio/Technology* 9:86, 1991; Palmiter et al., *Cell* 41:343, 1985; Kraemer et al., "Genetic Manipulation of the Early Mammalian Embryo," Cold Spring Harbor Press, Cold Spring Harbor, NY, 1985; Hammer et al., *Nature* 315:680, 1985; Purcel et al., *Science*, 244:1281, 1986; Wagner et al., U.S. Patent No. 5,175,385; and Krimpenfort et al., U.S. Patent No. 5,175,384 (the latter two publications are hereby incorporated by reference).

<u>Anti-Tango-78, Tango-79, or Tango-81 Antibodies</u>

Human Tango-78, Tango-79, and Tango-81 polypeptides (or immunogenic fragments or analogs) can be used to raise antibodies useful in the invention; such polypeptides can be produced by recombinant techniques or synthesized (see, for example, "Solid Phase Peptide Synthesis," *supra*; Ausubel et al., *supra*). In general, the peptides can be coupled to a carrier protein, such as KLH, as described in Ausubel et al., *supra*, mixed with an adjuvant, and injected into a host mammal. Antibodies can be purified by peptide antigen affinity chromatography.

In particular, various host animals can be immunized by injection with a Tango-78, Tango-79, or Tango-81 polypeptide. Host animals include rabbits, mice, guinea pigs, and rats. Various adjuvants that can be used to increase the immunological response depend on the host species and include Freund's adjuvant (complete and incomplete), mineral gels such as aluminum hydroxide,

surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Potentially useful human adjuvants include BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Polyclonal antibodies are heterogeneous populations of antibody molecules that are contained in the sera of the immunized animals.

Antibodies within the invention therefore include polyclonal antibodies and, in addition, monoclonal antibodies, humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, and molecules produced using a Fab expression library.

Monoclonal antibodies, which are homogeneous populations of antibodies to a particular antigen, can be prepared using the Tango-78, Tango-79, or Tango-81 polypeptides described above and standard hybridoma technology (see, for example, Kohler et al., *Nature* 256:495, 1975; Kohler et al., *Eur. J. Immunol.* 6:511, 1976; Kohler et al., *Eur. J. Immunol.* 6:292, 1976; Hammerling et al., "Monoclonal Antibodies and T Cell Hybridomas," Elsevier, NY, 1981; Ausubel et al., *supra*).

In particular, monoclonal antibodies can be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture such as described in Kohler et al., *Nature* 256:495, 1975, and U.S. Patent No. 4,376,110; the human B-cell hybridoma technique (Kosbor et al., *Immunology Today* 4:72, 1983; Cole et al., *Proc. Natl. Acad. Sci. USA* 80:2026, 1983), and the EBV-hybridoma technique (Cole et al., "Monoclonal Antibodies and Cancer Therapy," Alan R. Liss, Inc., pp. 77-96, 1983). Such antibodies can be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD and any subclass thereof. The hybridoma producing the mAb of this invention may be cultivated *in vitro* or *in vivo*. The ability to produce high titers of mAbs

in vivo makes this a particularly useful method of production.

Once produced, polyclonal or monoclonal antibodies are tested for specific Tango-78, Tango-79, or Tango-81 recognition by Western blot or immunoprecipitation analysis by standard methods, e.g., as described in Ausubel et al., *supra*. Antibodies that specifically recognize and bind to Tango-78, Tango-79, or Tango-81 are useful in the invention. For example, such antibodies can be used in an immunoassay to monitor the level of Tango-78, Tango-79, or Tango-81 produced by a mammal (for example, to determine the amount or subcellular location of Tango-78, Tango-79, or Tango-81).

Preferably, antibodies of the invention are produced using fragments of the Tango-78, Tango-79, or Tango-81 protein which lie outside highly conserved regions and appear likely to be antigenic, by criteria such as high frequency of charged residues. In one specific example, such fragments are generated by standard techniques of PCR, and are then cloned into the pGEX expression vector (Ausubel et al., *supra*). Fusion proteins are expressed in *E. coli* and purified using a glutathione agarose affinity matrix as described in Ausubel, et al., *supra*.

In some cases it may be desirable to minimize the potential problems of low affinity or specificity of antisera. In such circumstances, two or three fusions can be generated for each protein, and each fusion can be injected into at least two rabbits. Antisera can be raised by injections in a series, preferably including at least three booster injections.

Antisera may also checked for its ability to immunoprecipitate recombinant Tango-78, Tango-79, and Tango-81 proteins or control proteins, such as glucocorticoid receptor, CAT, or luciferase.

The antibodies can be used, for example, in the detection of the Tango-78, Tango-79, or Tango-81 in a biological sample as part of a diagnostic assay. Antibodies also can be used in a screening assay to measure the effect of a candidate compound on expression or localization of Tango-78, Tango-79, or Tango-81. Additionally, such antibodies can be used in conjunction with the gene therapy techniques described to, for example, evaluate normal and/or genetically engineered Tango-78, Tango-79, and Tango-81-expressing cells prior to their introduction into the patient.

In addition, techniques developed for the production of "chimeric antibodies" (Morrison et al., *Proc. Natl. Acad. Sci. USA*, 81:6851, 1984; Neuberger et al., *Nature*, 312:604, 1984; Takeda et al., *Nature*, 314:452, 1984) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent Nos. 4,946,778, 4,946,778, and 4,704,692) can be adapted to produce single chain antibodies against a Tango-78, Tango-79, or Tango-81 or polypeptide. Single chain antibodies are formed by linking the heavy and light chain fragments of the Fv region via an amino acid bridge, resulting in a single chain polypeptide.

Antibody fragments that recognize and bind to specific epitopes can be generated by known techniques. For example, such fragments include but are not limited to $F(ab')_2$ fragments that can be produced by pepsin

digestion of the antibody molecule, and Fab fragments that can be generated by reducing the disulfide bridges of $F(ab')_2$ fragments. Alternatively, Fab expression libraries can be constructed (Huse et al., *Science*, 246:1275, 1989) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity.

Antibodies to Tango-78, Tango-79, or Tango-81 can, in turn, be used to generate anti-idiotype antibodies that resemble a portion of the protein using techniques well known to those skilled in the art (see, e.g., Greenspan et al., *FASEB J.* 7:437, 1993; Nissinoff, *J. Immunol.* 147:2429, 1991). For example, antibodies that bind to the protein and competitively inhibit the binding of a binding partner of the protein can be used to generate anti-idiotypes that resemble a binding partner binding domain of the protein and, therefore, bind and neutralize a binding partner of the protein. Such neutralizing anti-idiotypic antibodies or Fab fragments of such anti-idiotypic antibodies can be used in therapeutic regimens.

Antibodies can be humanized by methods known in the art. For example, monoclonal antibodies with a desired binding specificity can be commercially humanized (Scotgene, Scotland; Oxford Molecular, Palo Alto, CA). Fully human antibodies, such as those expressed in transgenic animals are also features of the invention (Green et al., *Nature Genetics* 7:13-21, 1994; see also U.S. Patents 5,545,806 and 5,569,825, both of which are hereby incorporated by reference).

The methods described herein in which anti-Tango-78, Tango-79, or Tango-81 antibodies are employed may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one specific Tango-78, Tango-79, or Tango-81 antibody reagent described herein, which may be conveniently used, for example, in

clinical settings, to diagnose patients exhibiting symptoms disorders associated with abberent expression of Tango-78, Tango-79, or Tango-81.

<u>Antisense Nucleic Acids</u>

Treatment regimes based on an "antisense" approach involve the design of oligonucleotides (either DNA or RNA) that are complementary to Tango-78, Tango-79, or Tango-81 mRNA. These oligonucleotides bind to the complementary Tango-78, Tango-79, or Tango-81 mRNA transcripts and prevent translation. Absolute complementarity, although preferred, is not required. A sequence "complementary" to a portion of an RNA, as referred to herein, means a sequence having sufficient complementarily to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarily and the length of the antisense nucleic acid. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

Oligonucleotides that are complementary to the 5′ end of the message, e.g., the 5′ untranslated sequence up to and including the AUG initiation codon, should work most efficiently at inhibiting translation. However, sequences complementary to the 3′ untranslated sequences of mRNAs recently have been shown to be effective at inhibiting translation of mRNAs as well (Wagner, *Nature* 372:333, 1984). Thus, oligonucleotides complementary to either the 5′ or 3′ non-translated, non-coding regions of

the gene could be used in an antisense approach to inhibit translation of endogenous mRNA. Oligonucleotides complementary to the 5′ untranslated region of the mRNA should include the complement of the AUG start codon.

Antisense oligonucleotides complementary to mRNA coding regions are less efficient inhibitors of translation but could be used in accordance with the invention. Whether designed to hybridize to the 5′, 3′, or coding region of an mRNA, antisense nucleic acids should be at least six nucleotides in length, and are preferably oligonucleotides ranging from 6 to about 50 nucleotides in length. In specific aspects the oligonucleotide is at least 10 nucleotides, at least 17 nucleotides, at least 25 nucleotides, or at least 50 nucleotides.

Regardless of the choice of target sequence, it is preferred that *in vitro* studies are first performed to quantitate the ability of the antisense oligonucleotide to inhibit gene expression. It is preferred that these studies utilize controls that distinguish between antisense gene inhibition and nonspecific biological effects of oligonucleotides. It is also preferred that these studies compare levels of the target RNA or protein with that of an internal control RNA or protein. Additionally, it is envisioned that results obtained using the antisense oligonucleotide are compared with those obtained using a control oligonucleotide. It is preferred that the control oligonucleotide is of approximately the same length as the test oligonucleotide and that the nucleotide sequence of the oligonucleotide differs from the antisense sequence no more than is necessary to prevent specific hybridization to the target sequence.

The oligonucleotides can be DNA or RNA or chimeric mixtures or derivatives or modified versions thereof,

single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (as described, e.g., in Letsinger et al., *Proc. Natl. Acad. Sci. USA* 86:6553, 1989; Lemaitre et al., *Proc. Natl. Acad. Sci. USA* 84:648, 1987; PCT Publication No. WO 88/09810) or the blood-brain barrier (see, for example, PCT Publication No. WO 89/10134), or hybridization-triggered cleavage agents (see, for example, Krol et al., *BioTechniques* 6:958, 1988), or intercalating agents (see, for example, Zon, *Pharm. Res.* 5:539, 1988). To this end, the oligonucleotide can be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent.

The antisense oligonucleotide may comprise at least one modified base moiety which is selected from the group including, but not limited to, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethyl-aminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic

acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-theouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 2-(3-amino-3-N-2-carboxypropl) uracil, (acp3)w, and 2,6-diaminopurine.

The antisense oligonucleotide may also comprise at least one modified sugar moiety selected from the group including, but not limited to, arabinose, 2-fluoroarabinose, xylulose, and hexose.

In yet another embodiment, the antisense oligonucleotide comprises at least one modified phosphate backbone selected from the group consisting of a phosphorothioate, a phosphorodithioate, a phosphoramidothioate, a phosphoramidate, a phosphordiamidate, a methylphosphonate, an alkyl phosphotriester, and a formacetal, or an analog of any of these backbones.

In yet another embodiment, the antisense oligonucleotide is an α-anomeric oligonucleotide. An α-anomeric oligonucleotide forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gautier et al., *Nucl. Acids. Res.* 15:6625, 1987). The oligonucleotide is a 2'-0-methylribonucleotide (Inoue et al., *Nucl. Acids Res.* 15:6131, 1987), or a chimeric RNA-DNA analog (Inoue et al., *FEBS Lett.* 215:327, 1987).

Antisense oligonucleotides of the invention can be synthesized by standard methods known in the art, e.g., by use of an automated DNA synthesizer (such as are commercially available from Biosearch, Applied Biosystems, etc.). As examples, phosphorothioate oligonucleotides can be synthesized by the method of Stein et al. (*Nucl. Acids Res.* 16:3209, 1988), and methylphosphonate oligonucleotides can be prepared by use

of controlled pore glass polymer supports (Sarin et al., *Proc. Natl. Acad. Sci. USA* 85:7448, 1988).

The antisense molecules should be delivered to cells that express Tango-78, Tango-79, or Tango-81 *in vivo*. A number of methods have been developed for delivering antisense DNA or RNA to cells; e.g., antisense molecules can be injected directly into the tissue site, or modified antisense molecules, designed to target the desired cells (e.g., antisense linked to peptides or antibodies that specifically bind receptors or antigens expressed on the target cell surface) can be administered systemically.

However, it is often difficult to achieve intracellular concentrations of the antisense molecule sufficient to suppress translation of endogenous mRNAs. Therefore, a preferred approach uses a recombinant DNA construct in which the antisense oligonucleotide is placed under the control of a strong *pol* III or *pol* II promoter. The use of such a construct to transfect target cells in the patient will result in the transcription of sufficient amounts of single stranded RNAs that will form complementary base pairs with the endogenous Tango-78, Tango-79, or Tango-81 transcripts and thereby prevent translation of the endogenous mRNA. For example, a vector can be introduced *in vivo* such that it is taken up by a cell and directs the transcription of an antisense RNA. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA.

Such vectors can be constructed by recombinant DNA technology methods standard in the art. Vectors can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression of the sequence encoding the antisense RNA can be by any promoter known in the art to act in mammalian,

preferably human cells. Such promoters can be inducible or constitutive. Such promoters include, but are not limited to: the SV40 early promoter region (Bernoist et al., *Nature* 290:304, 1981); the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto et al., *Cell* 22:787-797, 1988); the herpes thymidine kinase promoter (Wagner et al., *Proc. Natl. Acad. Sci. USA* 78:1441, 1981); or the regulatory sequences of the metallothionein gene (Brinster et al., *Nature* 296:39, 1988).

Ribozymes

Ribozyme molecules designed to catalytically cleave Tango-78, Tango-79, or Tango-81 mRNA transcripts can be used to prevent translation of Tango-78, Tango-79, or Tango-81 mRNA. (see, e.g., PCT Publication WO 90/11364; Saraver et al., *Science* 247:1222, 1990). While various ribozymes that cleave mRNA at site-specific recognition sequences can be used to destroy Tango-78, Tango-79, or Tango-81 mRNAs, the use of hammerhead ribozymes is preferred. Hammerhead ribozymes cleave mRNAs at locations dictated by flanking regions that form complementary base pairs with the target mRNA. The sole requirement is that the target mRNA have the following sequence of two bases: 5'-UG-3'. The construction and production of hammerhead ribozymes is well known in the art (Haseloff et al., *Nature* 334:585, 1988). There are numerous examples of potential hammerhead ribozyme cleavage sites within the nucleotide sequence of human Tango-78, Tango-79, and Tango-81 cDNA. Preferably, the ribozyme is engineered so that the cleavage recognition site is located near the 5' end of the mRNA, i.e., to increase efficiency and minimize the intracellular accumulation of non-functional mRNA transcripts.

The ribozymes of the present invention also include RNA endoribonucleases (hereinafter "Cech-type

ribozymes"), such as the one that occurs naturally in *Tetrahymena Thermophila* (known as the IVS or L-19 IVS RNA), and which has been extensively described by Cech and his collaborators (Zaug et al., *Science* 224:574, 1984; Zaug et al., *Science*, 231:470, 1986; Zug et al., *Nature* 324:429, 1986; PCT Application No. WO 88/04300; and Been et al., *Cell* 47:207, 1986). The Cech-type ribozymes have an eight base-pair sequence that hybridizes to a target RNA sequence, whereafter cleavage of the target RNA takes place. The invention encompasses those Cech-type ribozymes that target eight base-pair active site sequences present in Tango-78, Tango-79, and Tango-81.

As in the antisense approach, the ribozymes can be composed of modified oligonucleotides (e.g., for improved stability, targeting, etc.), and should be delivered to cells which express Tango-78, Tango-79, or Tango-81 *in vivo*. A preferred method of delivery involves using a DNA construct "encoding" the ribozyme under the control of a strong constitutive *pol* III or *pol* II promoter, so that transfected cells will produce sufficient quantities of the ribozyme to destroy endogenous messages and inhibit translation. Because ribozymes, unlike antisense molecules, are catalytic, a lower intracellular concentration is required for efficiency.

<u>Other Methods for Reducing Tango-78, Tango-79, and Tango-81 Expression</u>

Endogenous Tango-78, Tango-79, and Tango-81 gene expression can also be reduced by inactivating the gene or its promoter using targeted homologous recombination (see, e.g., U.S. Patent No. 5,464,764). For example, a mutant, non-functional Tango-78, Tango-79, or Tango-81 gene (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous Tango-78, Tango-79, or Tango-81 gene (either the coding regions or regulatory

regions of the Tango-78, Tango-79, or Tango-81 gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express Tango-78, Tango-79, or Tango-81 *in vivo*. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the Tango-78, Tango-79, or Tango-81 gene. Such approaches are particularly suited for use in the agricultural field where modifications to ES (embryonic stem) cells can be used to generate animal offspring with an inactive Tango-78, Tango-79, or Tango-81. However, this approach can be adapted for use in humans, provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors.

Alternatively, endogenous Tango-78, Tango-79, or Tango-81 gene expression can be reduced by targeting deoxyribonucleotide sequences complementary to the regulatory region of the Tango-78, Tango-79, or Tango-81 gene (*i.e.*, the Tango-78, Tango-79, or Tango-81 promoter and/or enhancers) to form triple helical structures that prevent transcription of the Tango-78, Tango-79, or Tango-81 gene in target cells in the body (Helene *Anticancer Drug Res*. 6:569, 1981; Helene et al., *Ann. N.Y. Acad. Sci*. 660:27, 1992; and Maher, *Bioassays* 14:807, 1992).

<u>Detecting Proteins Associated with Tango-78, Tango-79, or Tango-81</u>

The invention also features polypeptides which interact with Tango-78, Tango-79, or Tango-81. Any method suitable for detecting protein-protein interactions may be employed for identifying transmembrane proteins, intracellular, or extracellular proteins that interact with Tango-78, Tango-79, or Tango-81. Among the traditional methods which may be employed are co-immunoprecipitation, cross-linking and

co-purification through gradients or chromatographic columns of cell lysates or proteins obtained from cell lysates and the use of Tango-78, Tango-79, or Tango-81 to identify proteins in the lysate that interact with Tango-78, Tango-79, or Tango-81. For these assays, the Tango-78, Tango-79, or Tango-81 polypetide can be full length Tango-78, Tango-79, or Tango-81, a soluble extracellular domain of Tango-78, Tango-79, and Tango-81, or some other suitable Tango-78, Tango-79, or Tango-81 polypeptide. Once isolated, such an interacting protein can be identified and cloned and then used, in conjunction with standard techniques, to identify proteins with which it interacts. For example, at least a portion of the amino acid sequence of a protein which interacts with the Tango-78, Tango-79, or Tango-81 can be ascertained using techniques well known to those of skill in the art, such as via the Edman degradation technique. The amino acid sequence obtained may be used as a guide for the generation of oligonucleotide mixtures that can be used to screen for gene sequences encoding the interacting protein. Screening may be accomplished, for example, by standard hybridization or PCR techniques. Techniques for the generation of oligonucleotide mixtures and the screening are well-known. (Ausubel, *supra;* and "PCR Protocols: A Guide to Methods and Applications," Innis et al., eds. Academic Press, Inc., NY, 1990).

Additionally, methods may be employed which result directly in the identification of genes which encode proteins which interact with Tango-78, Tango-79, or Tango-81. These methods include, for example, screening expression libraries, in a manner similar to the well known technique of antibody probing of λgt11 libraries, using labeled Tango-78, Tango-79, or Tango-81 polypeptide or a Tango-78, Tango-79, or Tango-81 fusion protein, e.g., a Tango-78, Tango-79, or Tango-81 polypeptide or

domain fused to a marker such as an enzyme, fluorescent dye, a luminescent protein, or to an IgFc domain.

There are also methods which are capable of detecting protein interaction. A method which detects protein interactions *in vivo* is the two-hybrid system (Chien et al., *Proc. Natl. Acad. Sci. USA*, 88:9578, 1991). A kit for practicing this method is available from Clontech (Palo Alto, CA).

Identification of a Tango-78, Tango-79, or Tango-81 Receptor

Receptors of Tango-78, Tango-79, or Tango-81 can be identified as follows. First cells or tissues which bind Tango-78, Tango-79, or Tango-81 are identified. An expression library is prepared using mRNA isolated from Tango-78, Tango-79, or Tango-81 binding cells. The expression library is used to transfect; eulcaryatic cells, e.g., CHO cells. Detectably labelled Tango-78, Tango-79, or Tango-81 is used to identify clones which bind Tango-78, Tango-79, or Tango-81. These clones are isolated and purified. The expression plasmid is then isolated from the Tango-78, Tango-79, or Tango-81-binding clones. These expression plasmids will encode putative Tango-78, Tango-79, or Tango-81 receptors.

Cells or tissues bearing a Tango-78, Tango-79, or Tango-81 receptor can be identified by exposing detectably labelled Tango-78, Tango-79, or Tango-81 to various cells lines and tissues. Alternatively a microphysiometer can be used to determine whether a selected cells responds to the presence of a cell receptor ligand (McConnel et al., *Science* 257:1906, 1992).

Compounds which bind Tango-78, Tango-79, or Tango-81

Compounds which bind Tango-78, Tango-79, or Tango-81 can be identified using any standard binding assay. For example, candidate compounds can be bound to a solid support. Tango-78, Tango-79, or Tango-81 is then exposed to the immobilized compound and binding is measured (European Patent Application 84/03564).

Effective Dose

Toxicity and therapeutic efficacy of the polypeptides of the invention and the compounds that modulate their expression or activity can be determined by standard pharmaceutical procedures, using either cells in culture or experimental animals to determine the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Polypeptides or other compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays.

A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (that is, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

Formulations and Use

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers or excipients.

Thus, the compounds and their physiologically acceptable salts and solvates may be formulated for administration by inhalation or insufflation (either through the mouth or the nose) or oral, buccal, parenteral or rectal administration.

For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (for example, pregelatinised maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (for example, lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (for example, magnesium stearate, talc or silica); disintegrants (for example, potato starch or sodium starch glycolate); or wetting agents (for example, sodium lauryl sulphate). The tablets may be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by

conventional means with pharmaceutically acceptable additives such as suspending agents (for example, sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (for example, lecithin or acacia); non-aqueous vehicles (for example, almond oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (for example, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations may also contain buffer salts, flavoring, coloring and sweetening agents as appropriate. Preparations for oral administration may be suitably formulated to give controlled release of the active compound.

For buccal administration the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, for example, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, for example, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, for example, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, for example, in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous

vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, for example, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, for example, containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may for example comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration.

The therapeutic compositions of the invention can also contain a carrier or excipient, many of which are known to skilled artisans. Excipients which can be used include buffers (for example, citrate buffer, phosphate buffer, acetate buffer, and bicarbonate buffer), amino acids, urea, alcohols, ascorbic acid, phospholipids, proteins (for example, serum albumin), EDTA, sodium chloride, liposomes, mannitol, sorbitol, and glycerol. The nucleic acids, polypeptides, antibodies, or modulatory compounds of the invention can be administered

by any standard route of administration. For example, administration can be parenteral, intravenous, subcutaneous, intramuscular, intracranial, intraorbital, opthalmic, intraventricular, intracapsular, intraspinal, intracisternal, intraperitoneal, transmucosal, or oral. The modulatory compound can be formulated in various ways, according to the corresponding route of administration. For example, liquid solutions can be made for ingestion or injection; gels or powders can be made for ingestion, inhalation, or topical application. Methods for making such formulations are well known and can be found in, for example, "Remington's Pharmaceutical Sciences." It is expected that the preferred route of administration will be intravenous.

Examples

Tango-78 cDNA (SEQ ID NO:1; FIG. 1) was isolated from a human bone marrow cDNA library (Clonetech; Palo Alto, CA). This Tango-78 cDNA encodes a 169 amino acid portion of Tango-78, a novel protein (SEQ ID NO:2; FIG. 1) that is highly homologous to the murine nodal protein (Collignon et al., *Nature* 381:155, 1996).

The Tango-78 cDNA (SEQ ID NO:1; FIG. 1) described herein was isolated using the method described in U.S. Serial No. 08/752,307 (filed November 19, 1996), hereby incorporated by reference. Tango-78 protein (SEQ ID NO:2; FIG. 1) is highly homologous to murine nodal protein (Collignon et al., *supra*; FIG.

Tango-79 cDNA (SEQ ID NO:3; FIG. 2) was isolated from a human fetal brain library (Clonetech; Palo Alto, CA). This Tango-78 cDNA encodes a 615 amino acid protein (SEQ ID NO:4; FIG. 2) that is homologous to *Drosophila Melanogaster* slit protein (Taguchi et al., *Brain Res. Mol. Brain Res.* 35:31, 1996).

The Tango-79 cDNA (SEQ ID NO:3; FIG. 2) described herein was isolated using the method described in U.S. Serial No. 08/752,307 (filed November 19, 1996), hereby incorporated by reference. Tango-79 protein (SEQ ID NO:4; FIG. 2) is homologous to D45913 (leucine rich repeat protein) (FIG. 5). Northern blot analysis of Tango-79 mRNA show that an approximate 3.0 kB and an approximate 3.5 kB transpcript are expressed in the brain. Tango-79 function can be studied by overexpressing the protein in mouse brain.

Tango-81 cDNA was isolated from a human fetal brain library. This Tango-81 cDNA (SEQ ID NO:5; FIG. 3) encodes a 261 amino acid protein (SEQ ID NO:6; FIG. 3).

The Tango-81 cDNA described herein was isolated using the method described in U.S. Serial No. 08/752,307 (filed November 19, 1996), hereby incorporated by reference. Northern analysis of Tango-81 expression reveals that it is expressed in heart, brain, spleen, lung, liver, skeletal muscle, kidneys and testis (FIG. 6).

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the cDNA insert of the plasmid deposited with ATCC as Accession Number _______, the cDNA insert of the plasmid deposited with ATCC as Accession Number _______, the cDNA insert of the plasmid deposited with ATCC as Accession Number _______or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the cDNA insert of the plasmid deposited with ATCC as Accession Number _______ , the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, the cDNA insert of the plasmid deposited with ATCC as Accession Number _________ or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _________, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _________, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _______;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO: 4, SEQ ID NO:6, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the polypeptide encoded by the cDNA insert of the plasmid

deposited with ATCC as Accession Number ___________, the polypeptide encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ___________, or the polypeptide encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ___________; and

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or a complement thereof under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, the cDNA insert of the plasmid deposited with ATCC as Accession Number _______, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____________, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____________, or

an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number _______,

wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or a complement thereof under stringent conditions; and

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or a complement thereof.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________.

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________,

or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ___________;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ________, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number __________, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number __________, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or a complement thereof under stringent conditions;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. The isolated polypeptide of claim 12 comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____, an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number ____.

14. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

15. The method of claim 14, wherein the compound which binds to the polypeptide is an antibody.

16. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

17. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

18. The method of claim 17, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

19. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

20. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

21. The method of claim 20, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;

c) detection of binding using an assay for Tango-72-mediated signal transduction.

22. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a

sufficient concentration to modulate the activity of the polypeptide.

23. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
CCTCGAGGGGAGGGGGAGATGCAACTCATCACATTTTTACTGACTGTCCTCTGGCTGTGCAAGTTATCTTGGAAGGGGG
ACTGGAAGAACAGTAATTCGGAGTCTGGGCTTGGCAGTTGGGCAAATCCAGGTTTACTCTTGGCTCTGCCACCTTCCAA
GAATGACACCTTGGTCAGATCTTTTAACCACACTGAGCCTCAGTTTTCCTCATCTCTAAAAGGGACTCGAAAATCTTAC
CAACTCATAGAGTTGGGGTGAGAATTCGAAGGTAATTCTATATAAGGTAAGGCCTCCAGCAAGAGCTATGGTGGTTGTG
ACACTGACTGAGGCTGGGGGAGGCCCTCACTCACCCTCCTTCCTTCTTGGTTTTCTCCTACCCAGATGTGGCAGTGGAT
GGGCAGAACTGGACGTTTGCTTTTGACTTCTCCTTCCTGAGCCAACAAGAGGATCTGGCATGGGCTGAGCTCCGGCTGC
AGCTGTCCAGCCCTGTGGACCTCCCCACTGAGGGCTCACTTGCCATTGAGATTTTCCACCAGCCAAAGCCCGACACAGA
                                    M   D   L   F   T   V   T   L   S   Q   V    11
GCAGGCTTCAGACAGCTGCTTAGAGCGGTTTCAG ATG GAC CTA TTC ACT GTC ACT TTG TCC CAG GTC   33

 T   F   S   L   G   S   M   V   L   E   V   T   R   P   L   S   K   W   L   K    31
ACC TTT TCC TTG GGC AGC ATG GTT TTG GAG GTG ACC AGG CCT CTC TCC AAG TGG CTG AAG   93

 R   P   G   A   L   E   K   Q   M   S   R   V   A   G   E   C   W   P   R   P    51
CGC CCT GGG GCC CTG GAG AAG CAG ATG TCC AGG GTA GCT GGA GAG TGC TGG CCG CGG CCC  153

 P   T   P   P   A   T   N   V   L   L   M   L   Y   S   N   L   S   Q   E   Q    71
CCC ACA CCG CCT GCC ACC AAT GTG CTC CTT ATG CTC TAC TCC AAC CTC TCG CAG GAG CAG  213

 R   Q   L   G   G   S   T   L   L   W   E   A   E   S   S   W   R   A   Q   E    91
AGG CAG CTG GGT GGG TCC ACC TTG CTG TGG GAA GCC GAG AGC TCC TGG CGG GCC CAG GAG  273

 G   Q   L   S   W   E   W   G   K   R   H   R   R   H   H   L   P   D   R   S   111
GGA CAG CTG TCC TGG GAG TGG GGC AAG AGG CAC CGT CGA CAT CAC TTG CCA GAC AGA AGT  333

 Q   L   C   R   K   V   K   F   Q   V   D   F   N   L   I   G   W   G   S   W   131
CAA CTG TGT CGG AAG GTC AAG TTC CAG GTG GAC TTC AAC CTG ATC GGA TGG GGC TCC TGG  393

 I   I   Y   P   K   Q   Y   N   A   Y   R   C   E   G   E   C   P   N   P   V   151
ATC ATC TAC CCC AAG CAG TAC AAC GCC TAT CGC TGT GAG GGC GAG TGT CCT AAT CCT GTT  453

 G   E   E   F   H   P   T   N   H   A   Y   I   Q   V   G   C   Q   A           169
GGG GAG GAG TTT CAT CCG ACC AAC CAT GCA TAC ATC CAG GTG GGA TGC CAG GCG          507

                                                                                 508
T
```

FIG. 1

```
TTGGGACCCAGCAGGACACAGCAGCAGTCAGGTGCATGCTGGGACCGCGACGGACAGGCTGCCGCACCCCAGGCCCCCA   79
                                             M   L   A   G   G   V   R              7
GAGGCCAGTCTGTTTGCCTCCCAACGCCATCTGACCCAGGTGAGCAAGAGG ATG CTG GCG GGG GGC GTG AGG    151
 S   M   P   S   P   L   L   A   C   W   Q   P   I   L   L   L   V   L   G   S    27
AGC ATG CCC AGC CCC CTC CTG GCC TGC TGG CAG CCC ATC CTC CTG CTG GTG CTG GGC TCA   211
 V   L   S   G   S   A   T   G   C   P   P   R   C   E   C   S   A   Q   D   R    47
GTG CTG TCA GGC TCG GCC ACG GGC TGC CCG CCC CGC TGC GAG TGC TCC GCC CAG GAC CGC   271
 A   V   L   C   H   R   K   R   F   V   A   V   P   E   G   I   P   T   E   T    67
GCT GTG CTG TGC CAC CGC AAG CGC TTT GTG GCA GTC CCC GAG GGC ATC CCC ACC GAG ACG   331
 R   L   L   D   L   G   K   N   R   I   K   T   L   N   Q   D   E   F   A   S    87
CGC CTG CTG GAC CTA GGC AAG AAC CGC ATC AAA ACG CTC AAC CAG GAC GAG TTC GCC AGC   391
 F   P   H   L   E   E   L   E   L   N   E   N   I   V   S   A   V   E   P   G   107
TTC CCG CAC CTG GAG GAG CTG GAG CTC AAC GAG AAC ATC GTG AGC GCC GTG GAG CCC GGC   451
 A   F   N   N   L   F   N   L   R   T   L   G   L   R   S   N   R   L   K   L   127
GCC TTC AAC AAC CTC TTC AAC CTC CGG ACG CTG GGT CTC CGC AGC AAC CGC CTG AAG CTC   511
 I   P   L   G   V   F   T   G   L   S   N   L   T   K   L   D   T   R   E   N   147
ATC CCG CTA GGC GTC TTC ACT GGC CTC AGC AAC CTG ACC AAG CTG GAC ACG AGG GAG AAC   571
 K   I   V   I   L   L   D   Y   M   F   Q   D   L   Y   N   L   K   S   L   E   167
AAG ATC GTT ATC CTA CTG GAC TAC ATG TTT CAG GAC CTG TAC AAC CTC AAG TCA CTG GAG   631
 V   G   D   N   D   L   V   Y   I   S   H   R   A   F   S   G   L   N   S   L   187
GTT GGC GAC AAT GAC CTC GTC TAC ATC TCT CAC CGC GCC TTC AGC GGC CTC AAC AGC CTG   691
 E   Q   L   T   L   E   K   C   N   L   T   S   I   P   T   E   A   L   S   H   207
GAG CAG CTG ACT CTG GAG AAA TGC AAC CTG ACC TCC ATC CCC ACC GAG GCG CTG TCC CAC   751
 L   H   G   L   I   V   L   R   L   R   H   L   N   I   N   A   I   R   D   Y   227
CTG CAC GGC CTC ATC GTC CTG AGG CTC CGG CAC CTC AAC ATC AAT GCC ATC CGG GAC TAC   811
 S   F   K   R   L   Y   R   L   K   V   L   E   I   S   H   W   P   Y   L   D   247
TCC TTC AAG AGG CTG TAC CGA CTC AAG GTC TTG GAG ATC TCC CAC TGG CCC TAC TTG GAC   871
```

FIG. 2 (1 of 3)

```
 T   M   T   P   N   C   L   Y   G   L   N   L   T   S   L   S   I   T   H   C    267
ACC ATG ACA CCC AAC TGC CTC TAC GGC CTC AAC CTG ACG TCC CTG TCC ATC ACA CAC TGC   931

 N   L   T   A   V   P   Y   L   A   V   R   H   L   V   Y   L   R   F   L   N    287
AAT CTG ACC GCT GTG CCC TAC CTG GCC GTC CGC CAC CTA GTC TAT CTC CGC TTC CTC AAC   991

 L   S   Y   N   P   I   S   T   I   E   G   S   M   L   H   E   L   L   R   L    307
CTC TCC TAC AAC CCC ATC AGC ACC ATT GAG GGC TCC ATG TTG CAT GAG CTG CTC CGG CTG  1051

 Q   E   I   Q   L   V   G   G   Q   L   A   V   V   E   P   Y   A   F   R   G    327
CAG GAG ATC CAG CTG GTG GGC GGG CAG CTG GCC GTG GTG GAG CCC TAT GCC TTC CGC GGC  1111

 L   N   Y   L   R   V   L   N   V   S   G   N   Q   L   T   T   L   E   E   S    347
CTC AAC TAC CTG CGC GTG CTC AAT GTC TCT GGC AAC CAG CTG ACC ACA CTG GAG GAA TCA  1171

 V   F   H   S   V   G   N   L   E   T   L   I   L   D   S   N   P   L   A   C    367
GTC TTC CAC TCG GTG GGC AAC CTG GAG ACA CTC ATC CTG GAC TCC AAC CCG CTG GCC TGC  1231

 D   C   R   L   L   W   V   F   R   R   R   W   R   L   N   F   N   R   Q   Q    387
GAC TGT CGG CTC CTG TGG GTG TTC CGG CGC CGC TGG CGG CTC AAC TTC AAC CGG CAG CAG  1291

 P   T   C   A   T   P   E   F   V   Q   G   K   E   F   K   D   F   P   D   V    407
CCC ACG TGC GCC ACG CCC GAG TTT GTC CAG GGC AAG GAG TTC AAG GAC TTC CCT GAT GTG  1351

 L   L   P   N   Y   F   T   C   R   R   A   R   I   R   D   R   K   A   Q   Q    427
CTA CTG CCC AAC TAC TTC ACC TGC CGC CGC GCC CGC ATC CGG GAC CGC AAG GCC CAG CAG  1411

 V   F   V   D   E   G   H   T   V   Q   F   V   C   R   A   D   G   D   P   P    447
GTG TTT GTG GAC GAG GGC CAC ACG GTG CAG TTT GTG TGC CGG GCC GAT GGC GAC CCG CCG  1471

 P   A   I   L   W   L   S   P   R   K   H   L   V   S   A   K   S   N   G   R    467
CCC GCC ATC CTC TGG CTC TCA CCC CGA AAG CAC CTG GTC TCA GCC AAG AGC AAT GGG CGG  1531

 L   T   V   F   P   D   G   T   L   E   V   R   Y   A   Q   V   Q   D   N   G    487
CTC ACA GTC TTC CCT GAT GGC ACG CTG GAG GTG CGC TAC GCC CAG GTA CAG GAC AAC GGC  1591

 T   Y   L   C   I   A   A   N   A   G   G   N   D   S   M   P   A   H   L   H    507
ACG TAC CTG TGC ATC GCG GCC AAC GCG GGC GGC AAC GAC TCC ATG CCC GCC CAC CTG CAT  1651

 V   R   S   Y   S   P   D   W   P   H   Q   P   N   K   T   F   A   F   I   S    527
GTG CGC AGC TAC TCG CCC GAC TGG CCC CAT CAG CCC AAC AAG ACC TTC GCT TTC ATC TCC  1711
```

FIG. 2 (2 of 3)

```
 N   Q   P   G   E   G   E   A   N   S   T   R   A   T   V   P   F   P   F   D    547
AAC CAG CCG GGC GAG GGA GAG GCC AAC AGC ACC CGC GCC ACT GTG CCT TTC CCC TTC GAC  1771

 I   K   T   L   I   I   A   T   T   M   G   F   I   S   F   L   G   V   V   L    567
ATC AAG ACC CTC ATC ATC GCC ACC ACC ATG GGC TTC ATC TCT TTC CTG GGC GTC GTC CTC  1831

 F   C   L   V   L   L   F   L   W   S   R   G   K   G   N   T   K   H   N   I    587
TTC TGC CTG GTG CTG CTG TTT CTC TGG AGC CGG GGC AAG GGC AAC ACA AAG CAC AAC ATC  1891

 E   I   E   Y   V   P   R   K   S   D   A   G   I   S   S   A   D   A   P   R    607
GAG ATC GAG TAT GTG CCC CGA AAG TCG GAC GCA GGC ATC AGC TCC GCC GAC GCG CCC CGC  1951

 K   F   N   M   K   M   I   *                                                    615
AAG TTC AAC ATG AAG ATG ATA TGA                                                  1975

GGCCGGGGCGGGGGGCAGGGACCCCCGGGCGGCCGGGCAGGGGAAGGGGCCTGGCCGCCACCTGCTCACTCTCCAGTCC  2054

TTCCCACCTCCTCCCTACCCTTCTACACACGTTCTCTTTCTCCCCTCCCGCCTCCGTCCCCTGCTGCCCCCCGCCAGCC  2133

CTCACCACCTGCCCTCCTTCTACCAGGACCTCAGAAGCCCAGACCTGGGGACCCCACCTACACAGGGGCATTGACAGAC  2212

TGGAGTTTAAAGCCGACGAACCGACACGCGGCAGAGTCAATAATTCAATAAAAAAGTTACGAACTTTCTCTGTAACTTG  2291

GGTTTCAATAATTATGGATTTTTATGAAAACTTGAAATAATAAAAAAAAAAAAAAAAAAG                    2351
```

FIG. 2 (3 of 3)

```
                                                         M   A   G   S   P      5
GAATTCGGCACGAGGCCAGCCAGTCCGCCSGYMCGRRGCCCGGCTCGCTGGGGCAGC ATG GCG GGG TCG CCG   72

 L   L   W   G   P   R   A   G   G   V   G   L   L   V   L   L   L   L   G   L    25
CTG CTC TGG GGG CCG CGG GCC GGG GGC GTC GGC CTT TTG GTG CTG CTG CTG CTC GGC CTG  132

 F   R   P   P   P   A   L   C   A   R   P   V   K   E   P   R   G   L   S   A    45
TTT CGG CCG CCC CCC GCG CTC TGC GCG CGG CCG GTA AAG GAG CCC CGC GGC CTA AGC GCA  192

 A   S   P   P   L   A   E   T   G   A   P   R   R   F   R   R   S   V   P   R    65
GCG TCT CCG CCC TTG GCT GAG ACT GGC GCT CCT CGC CGC TTC CGG CGG TCA GTG CCC CGA  252

 G   E   A   A   G   A   V   Q   E   L   A   R   A   L   A   H   L   L   E   A    85
GGT GAG GCG GCG GGG GCG GTG CAG GAG CTG GCG CGG GCG CTG GCG CAT CTG CTG GAG GCC  312

 E   R   Q   E   R   A   R   A   E   A   Q   E   A   E   D   Q   Q   A   R   V   105
GAA CGT CAG GAG CGG GCG CGG GCC GAG GCG CAG GAG GCT GAG GAT CAG CAG GCG CGC GTC  372

 L   A   Q   L   L   R   V   W   G   A   P   R   N   S   D   P   A   L   G   L   125
CTG GCG CAG CTG CTG CGC GTC TGG GGC GCC CCC CGC AAC TCT GAT CCG GCT CTG GGC TTG  432

 D   D   D   P   D   A   P   A   A   Q   L   A   R   A   L   L   R   A   R   L   145
GAC GAC GAC CCC GAC GCG CCT GCA GCG CAG CTC GCT CGC GCT CTG CTC CGC GCC CGC CTT  492

 D   P   A   A   L   A   A   Q   L   V   P   A   P   V   P   A   A   A   L   R   165
GAC CCT GCC GCC CTA GCA GCC CAG CTT GTC CCC GCG CCC GTC CCC GCC GCG GCG CTC CGA  552

 P   R   P   P   V   Y   D   D   G   P   A   G   P   D   A   E   E   A   G   D   185
CCC CGG CCC CCG GTC TAC GAC GAC GGC CCC GCG GGC CCG GAT GCT GAG GAG GCA GGC GAC  612

 E   T   P   D   V   D   P   E   L   L   R   Y   L   L   G   R   I   L   A   G   205
GAG ACA CCC GAC GTG GAC CCC GAG CTG TTG AGG TAC TTG CTG GGA CGG ATT CTT GCG GGA  672

 S   A   D   S   E   G   V   A   A   P   R   R   L   R   R   A   A   D   H   D   225
AGC GCG GAC TCC GAG GGG GTG GCA GCC CCG CGC CGC CTC CGC CGT GCC GCC GAC CAC GAT  732

 V   G   S   E   L   P   P   E   G   V   L   G   A   L   L   R   V   K   R   L   245
GTG GGC TCT GAG CTG CCC CCT GAG GGC GTG CTG GGG GCG CTG CTG CGT GTG AAA CGC CTA  792

 E   T   P   A   P   Q   V   P   A   R   R   L   L   P   P   *                   261
GAG ACC CCG GCG CCC CAG GTG CCT GCA CGC CGC CTC TTG CCA CCC TGA                  840

GCACTGCCCGGATCCCGTGCACCCTGGGACCCAGAAGTGCCCCCGCCATCCCGCCACCAGGACTGCTCCCCGCCAGCAC  919

GTCCAGAGCAACTTACCCCGGCCAGCCAGCCCTCTCACCCGAGGATCCCTACCCCCTGGC                     979
```

FIG. 3

```
T78      1 ..................................MDLFTVTLSQVTFSLG  16
                                             |: |||. |||||. |
nodal  101 PMDIPTEGPLTIDIFHQAKGDPERDPADCLERIWMETFTVIPSQVTFASG 150

T78     17 SMVLEVTRPLSKWLKRPGALEKQMSRVAGECWPRPPTPP...ATNVLLML  63
           | |||||:||||||| | |||||:|. |:.||..| |||   |.. .|||
nodal  151 STVLEVTKPLSKWLKDPRALEKQVSSRAEKCWHQPYTPPVPVASTNVLML 200

T78     64 YSNLSQEQRQLGGSTLLWEAESSWRAQEGQLSWE...WGKRHRRHHLPDR 110
           ||| .||||||||.|||||||||||||||||| |   ||:|:||||||||
nodal  201 YSNRPQEQRQLGGATLLWEAESSWRAQEGQLSVERGGWGRRQRRHHLPDR 250

T78    111 SQLCRKVKFQVDFNLIGWGSWIIYPKQYNAYRCEGECPNPVGEEFHPTNH 160
           |||||:||||||||||||||||||||||||||||||||||||||||||||
nodal  251 SQLCRRVKFQVDFNLIGWGSWIIYPKQYNAYRCEGECPNPVGEEFHPTNH 300

T78    161 AYIQV...GCQA...................................... 169
           ||||     :|:
nodal  301 AYIQSLLKRYQPHRVPSTCCAPVKTKPLSMLYVDNGRVLLEHHKDMIVEE 350
```

FIG. 4

```
T79        1 MLAGGVRSMPSPLLACWQPILLLVLGSVLSGS..ATGCPPRCECSAQDR.  47
D45913     1 .....MARLSTGKAAC.QVVLGLLITSLTESSILTSECPQLCVCEIRPWF  44

T79       48 ..........AVLCHRKRFVAVPEGIPTETRLLDLGKNRIKTLNQDEFAS  87
D45913    45 TPQSTYREATTVDCNDLRLTRIPGNLSSDTQVLLLQSNNI..........  84

T79       88 FPHLEELELNENIVSAVEPGAFNNLFNLRTLGLRSNRLKLIPLGVFTGLS 137
D45913    85 ...............AKTVDELQQLFNLTELDFSQNNFTNIKEVGLANLT 119

T79      138 NLTKLDTRENKIVILLDYMFQDLYNLKSLEVGDNDLVYISHRAFSGLNSL 187
D45913   120 QLTTLHLEENQISEMTDYCLQDLSNLQELYINHNQISTISANAFSGLKNL 169

T79      188 EQLTLEKCNLTSIPTEALSHLHGLIVLRLRHLNINAIRDYSFKRLYRLKV 237
D45913   170 LRLHLNSNKLKVIDSRWFDSTPNLEILMIGENPVIGILDMNFRPLSNLRS 219

T79      238 LEISHWPYLDTMTPNCLYGLN.LTSLSITHCNLTAVPYLAVRHLVYLRFL 286
D45913   220 LVLAG.MYLTDVPGNALVGLDSLESLSFYDNKLIKVPQLALQKVPNLKFL 268

T79      287 NLSYNPISTIEGSMLHELLRLQEIQLVG.GQLAVVEPY............ 323
D45913   269 DLNKNPIHKIQEGDFKNMLRLKELGINNMGELVSVDRYALDNLPELTKLE 318

T79      324 .............AFRGLNYLRVLNVSGNQLTTLEESVFHSVGNLETLIL 360
D45913   319 ATNNPKLSYIHRLAFRSVPALESLMLNNNALNAVYQKTVESLPNLREISI 368

T79      361 DSNPLACDCRLLWVFRRRWRLNFNRQQPT.CATPEFVQGKEFKDFPDVLL 409
D45913   369 HSNPLRCDCVIHWINSNKTNIRFMEPLSMFCAMPPEYRGQQVK...EVLI 415

T79      410 PNYFT.CRRARIRDRKAQQVFVDEGHTVQFVCRADGDPPPAILWLSPRKH 458
D45913   416 QDSSEQCLPMISHDTFPNHLNMDIGTTLFLDCRAMAEPEPEIYWVTPIGN 465

T79      459 LVSAKS.NGRLTVFPDGTLEVRYAQVQDNGTYLCIAANAGGNDSMPAHLH 507
D45913   466 KITVETLSDKYKLSSEGTLEIANIQIEDSGRYTCVAQNVQGADTRVATIK 515

T79      508 V.................................RSYSPDWPHQ 518
D45913   516 VNGTLLDGAQVLKIYVKQTESHSILVSWKVNSNVMTSNLKWSSATMKIDN 565

T79      519 PNKTF..........................AFISNQPGEGEANSTRA 540
D45913   566 PHITYTARVPVDVHEYNLTHLQPSTDYEVCLTVSNIHQQTQKSCVNVTTK 615

T79      541 TVPFPFDIKTLIIATTMGFI..SFLGVVLFCLVLLFLWSRGKGNTKHNIE 588
D45913   616 TAAFALDISDHETSTALAAVMGSMFAVISLASIAIYIAKRFKRKNYHHSL 665

T79      589 IEYVPRKSDAGISSADAPRKFNMKMI........................ 614
D45913   666 KKYMQKTSSIPLNEL.YPPLINLWEADSDKDKDGSADTKPTQVDTSRSYY 714
```

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/16241

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/02, 21/04, 1/00, 17/00; C12Q 1/68; G01N 33/53
US CL : 536/23.1; 530/350, 387.1; 435/6, 7.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 530/350, 387.1; 435/6, 7.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
DIALOG: MEDLINE, BIOSIS, WPI, USPATFUL. author and terms (e.g. "TANGO" and protein) searched.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Database MEDLINE on STN, 98072332, SONNENFELD, ET AL., The Drosophila tango gene encodes a bHLH-PAS protein that is othologous to mammalian Arnt and controls CNS midline and tracheal development. Development, November 1997, Volume 124, Number 22, pages 4571-82, Abstract.	1-23

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 OCTOBER 1998	04 NOV 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer HEATHER BAKALYAR for Telephone No. (703) 308-0196